Tarski-style Axiomatic Truth Theories

Ali Enayat

Proof Theory Society Summer School (2025)
Ghent University

Lecture 1 + Lecture 2 + Lecture 3

Preamble

The subject matter of these lectures lie at the intersection of (1) Axiomatic Truth Theories, (2) Axiomatic Arithmetic, and (3) Axiomatic Set Theory. The following textbooks provide the foundations for each of these three subjects.

1. Axiomatic Truth Theories:
    - C. Cieśliński, **The Epistemic Lightness of Truth**. **Deflationism and its Logic**, Cambridge University Press, Cambridge, 2017.
    - V. Halbach, **Axiomatic Theories of Truth** (second ed), Cambridge University Press, 2015.
2. Axiomatic Arithmetic
    - P. Hájek and P. Pudlák, **Metamathematics of First-Order Arithmetic**, Springer, 1993.
    - R. Kaye, **Models of Peano Arithmetic**, Oxford University Press, 1991.
3. Axiomatic Set Theory
    - A. Levy, **Basic Set Theory**, Dover Publications, 2003.
    - K. Hrbacek and T. Jech, **Introduction to Set Theory** (third ed), 1999.

Tarskian Satisfaction, defined (1)

- Let $\mathcal{M} = (M, \cdots)$ be an $\mathcal{L}$-structure ($\mathcal{L}$ is the language/signature of $\mathcal{M}$); and φ be an $\mathcal{L}$-(first order) formula; we will assume that first order logic only uses the logical constants $\{\neg, \vee, \exists\}$.
- Also let α be an assignment for φ, i.e., $\alpha : \mathrm{FV}(\varphi) \to M$, where $\mathrm{FV}(\varphi)$ is the set of free variables of φ.
- The ternary relation $\boxed{\mathcal{M} \models \varphi[\alpha]}$ is defined by recursion on the complexity of φ as follows:
  1. $\mathcal{M} \models R(v_1, ..., v_k)[\alpha]$ iff $R^{\mathcal{M}}(\alpha(v_1), ..., \alpha(v_k))$, and more generally $\mathcal{M} \models R(t_1, ..., t_n)[\alpha]$ iff $R^{\mathcal{M}}(\hat{\alpha}(t_1), ..., \hat{\alpha}(t_n))$.

     Here $\hat{\alpha}$ is extension of α to terms whose free variables are in $\mathrm{Dom}(\alpha)$.
  2. $\mathcal{M} \models \neg\varphi[\alpha]$ iff $\mathcal{M} \not\models \varphi[\alpha]$.
  3. $\mathcal{M} \models (\varphi_1 \vee \varphi_2)[\alpha]$ iff $\mathcal{M} \models \varphi_1[\alpha_1]$ or $\mathcal{M} \models \varphi_2[\alpha_2]$, where $\alpha_i = \alpha \restriction \mathrm{FV}(\varphi_i)$.
  4. $\mathcal{M} \models \exists v\ \varphi[\alpha]$ iff there is some $m \in M$ such that $\mathcal{M} \models \varphi[\alpha^v_m]$.

α^v_m is the modification of α that sends the variable v to m but is otherwise the same as α

Lecture 1 + Lecture 2 + Lecture 3

Tarskian Satisfaction, defined (2)

- We can write a set-theoretic formula $\boxed{\mathsf{Sat}_{\mathcal{M}}(\varphi,\alpha)}$ such that within a sufficiently strong fragment of ZF, the following are provable:
  1. The recursive clauses on the previous page hold when $\mathcal{M} \models \varphi[\alpha]$ is replaced by $\mathsf{Sat}_{\mathcal{M}}(\varphi,\alpha)$.
  2. $\{(\varphi,\alpha) : \mathsf{Sat}_{\mathcal{M}}(\varphi,\alpha)\}$ forms a set, denoted $\boxed{\mathsf{sat}_{\mathcal{M}}}$.
- We will refer to $\mathsf{sat}_{\mathcal{M}}$ as the Tarskian satisfaction relation on $\mathcal{M}$.
- The theory of $\mathcal{M}$, denoted $\boxed{\mathsf{Th}(\mathcal{M})}$ is then defined as the set of $\mathcal{L}$-sentences σ such that $(\sigma,\varnothing) \in \mathsf{sat}_{\mathcal{M}}$.
- The following two well-known finitely axiomatized fragments of ZF are sufficiently strong for the above purposes:
  1. Zermelo set theory with the separation scheme limited to Δ_0-formulas.
  2. Kripke-Platek set theory with the $\in$-induction scheme limited to Σ_1-formulas.
- For finite structures $\mathcal{M}$, $\mathrm{I}\Delta_0 + \mathsf{Exp}$ is strong enough to construct $\mathsf{sat}_{\mathcal{M}}$.

Truth vs Satisfaction

- Given an $\mathcal{L}$-structure $\mathcal{M}$, let $\mathcal{L}_M$ be the result of enriching $\mathcal{L}$ with constant symbols $\dot{m}$ for each $m \in M$.
- The binary relation $\boxed{\mathcal{M} \models \sigma}$ is defined by recursion on the complexity of $\mathcal{L}_M$-sentences σ using the following clauses:
  1. $\mathcal{M} \models R(\dot{m}_1, ..., \dot{m}_k)$ iff $R^{\mathcal{M}}(m_1, ..., m_k)$, and more generally: $\mathcal{M} \models R(t_1, ..., t_k)$ iff $R^{\mathcal{M}}(t_1^{\mathcal{M}}, ..., t_k^{\mathcal{M}})$. Each t_i is a closed $\mathcal{L}$-term.
  2. $\mathcal{M} \models \neg\sigma$ iff $\mathcal{M} \not\models \sigma$.
  3. $\mathcal{M} \models \sigma_1 \vee \sigma_2$ iff $\mathcal{M} \models \sigma_1$ or $\mathcal{M} \models \sigma_2$.
  4. $\mathcal{M} \models \exists v\ \varphi(v)$ iff there is some $m \in M$ such that $\mathcal{M} \models \varphi(\dot{m}/v)$.
- Let $\boxed{\mathcal{M}^+}$ be expansion $(\mathcal{M}, m)_{m \in M}$ of $\mathcal{M}$.
- $\mathrm{Th}(\mathcal{M}^+)$ is referred to as the elementary diagram of $\mathcal{M}$ by model-theorists.
- In a sufficiently strong meta-theory, $\mathrm{sat}_{\mathcal{M}}$ and $\mathrm{Th}(\mathcal{M}^+)$ for a given structure can be defined from each other, but in the abstract axiomatic framework, satisfaction and truth have subtle differences.

Undefinability of Truth

- Let T be an $\mathcal{L}$-theory. Suppose $n \mapsto \underline{n}$ be a mapping of ω (natural numbers) into the set of closed $\mathcal{L}$-terms (terms with no free variables).
  Fix a 1-1 correspondence $\varphi \mapsto \#(\varphi)$ between $\text{Form}^1_{\mathcal{L}}$ and ω, and let $n \longmapsto \varphi_n$ be its inverse.

  The *diagonal function* $\delta : \omega \to \omega$ is given by $\delta(n) = \#(\varphi_n(\underline{n}))$.

  $f : \omega \to \omega$ is *T-definable* if there is an $\mathcal{L}$-formula $\theta(x, y)$ such that

$$\forall n \in \omega \;\; T \vdash \forall y \;\; \left[\theta(\underline{n}, y) \leftrightarrow y = \underline{f(n)}\right].$$

- **Theorem.** (Tarski Undefinability of Truth). *If T is a consistent $\mathcal{L}$-theory such that δ is T-definable, then there is no $\mathcal{L}$-formula $\theta(x)$ such that for all $\mathcal{L}$-sentences φ we have: $T \vdash \varphi \leftrightarrow \theta(\underline{\#(\varphi)})$.*

- **Corollary**. *If δ is* $\text{Th}(\mathcal{M})$*-definable, then* $\text{Th}(\mathcal{M})$ *is not parameter-free definable in $\mathcal{M}$.*

Peano Arithmetic

- PA (Peano Arithmetic) is the first order theory formulated in the language (signature) $\mathcal{L}_{\mathrm{PA}}$ consisting of 0, 1, $+$, $\times$, $<$, whose axioms consist of

$$\mathsf{Q} + \mathsf{Ind},$$

where Q (Robinson arithmetic) is finitely axiomatized, and

$$\mathsf{Ind} = \{\mathsf{Ind}_{\varphi(x,v)} : \varphi(x,v) \in \mathrm{Form}_{\mathrm{PA}}\},$$

where $\mathsf{Ind}_{\varphi(x,v)}$ is the following instance of induction:

$$\forall v \left[[(\varphi(0,v) \wedge \forall x(\varphi(x,v) \rightarrow \varphi(x+1,v)] \rightarrow \forall x\ \varphi(x,v) \right].$$

- Another common axiomatizations of PA is

$$\mathsf{PA}^- + \mathsf{Ind},$$

where PA^- is the theory of the non-negative parts of discrete ordered (a finitely axiomatized theory).

The truth theory $\mathrm{CT}^-[\mathrm{PA}]$

$\mathrm{CT}^-[\mathrm{PA}] = \mathrm{PA} + \mathrm{CT}^-$, where the axioms of CT^- are (1) – (5) below. The axioms are formulated in $\mathcal{L}_{\mathrm{PA}}(\mathrm{T}) := \mathcal{L}_{\mathrm{PA}} \cup \{\mathrm{T}(x)\}$.

1. $\forall x \, [\mathrm{T}(x) \to \mathrm{Sent}_{\mathrm{PA}}(x)]$
2. $\forall s, t \in \mathrm{ClTerm}_{\mathrm{PA}} \; [\mathrm{T}(s = t) \leftrightarrow s^\circ = t^\circ]$.
3. $\forall \varphi \in \mathrm{Sent}_{\mathrm{PA}} \, [\mathrm{T}(\neg\varphi) \leftrightarrow \neg \mathrm{T}(\varphi)]$.
4. $\forall \psi_1, \psi_2 \in \mathrm{Sent}_{\mathrm{PA}} \, [\mathrm{T}(\psi_1 \vee \psi_2) \leftrightarrow (\mathrm{T}(\psi_1) \vee \mathrm{T}(\psi_2))]$.
5. $\forall \psi(v) \in \mathrm{Form}^1_{\mathrm{PA}} \; \left[\mathrm{T}(\exists v \varphi(v)) \leftrightarrow \exists x \; \mathrm{T}\varphi(\dot{x}/v)\right]$.

$\mathrm{Sent}_{\mathrm{PA}}(x)$ expresses "x is (the code of) an $\mathcal{L}_{\mathrm{PA}}$-sentence";

$\mathrm{ClTerm}_{\mathrm{PA}}$ is the set of (codes of) closed terms of $\mathcal{L}_{\mathrm{PA}}$;

s° denotes the value of s;

$\mathrm{Form}^1_{\mathrm{PA}}(x)$ expresses "x is (the code of) an $\mathcal{L}_{\mathrm{PA}}$-formula with one free variable".

The truth theory CT[PA] and its fragments

- Let Ind(T) be the induction scheme of PA extended to $\mathcal{L}_{\mathsf{PA}}(\mathsf{T})$.
- CT[PA] $:= \mathsf{CT}^{-}[\mathsf{PA}] + \mathsf{Ind}(\mathsf{T})$,
- The intended model of CT[PA] is $(\omega, +, \cdot, \mathsf{Th}(\omega, +, \cdot))$.
- $\mathsf{TB}^{-} = \{\varphi \leftrightarrow T(\ulcorner\varphi\urcorner) : \varphi \in \mathsf{Sent}_{\mathsf{PA}}\}$.
- $\mathsf{TB} = \mathsf{TB}^{-} + \mathsf{Ind}(\mathsf{T})$
- $\mathsf{UTB}^{-} = \{\forall x[\varphi(x) \leftrightarrow \mathsf{T}(\ulcorner\varphi(\dot{x})\urcorner)] : \varphi(x) \in \mathsf{Form}^{1}_{\mathsf{PA}}\}$.
- $\mathsf{UTB} = \mathsf{UTB}^{-} + \mathsf{Ind}(\mathsf{T})$
- $\mathsf{CT}^{-} \vdash \mathsf{UTB}^{-} \vdash \mathsf{TB}^{-}$.
- $\mathsf{CT} \vdash \mathsf{UTB} \vdash \mathsf{UTB}$.

Base theories and Truth theories

- A base theory $\mathbb{B}$ is a first order theory with "enough coding for handling finite sequences of objects" (a sequential theory). For example:
  (1) $\mathbb{B} = \mathrm{PA}$ (Peano arithmetic).
  (2) $\mathbb{B} = \mathrm{ACA}_0$ (the predicative extension of PA).
  (3) $\mathbb{B} = \mathrm{ZF}$ (Zermelo-Fraenkel set theory).
  (4) $\mathbb{B} = \mathrm{GB}$ (the predicative extension of ZF).

  But not Robinson's Q!
- A truth theory over a base theory $\mathbb{B}$ is a theory of the form:

$$\mathrm{P}[\mathbb{B}] = \mathbb{B} \cup \mathrm{P},$$

  where P (for prawda = truth in Polish) is a set of "truth axioms" formulated in the language $\mathcal{L}_{\mathbb{B}} \cup \{\mathrm{T}(x)\}$, where the intended interpretation of $\mathrm{T}(x)$ is "x is the Gödel number of a true $\mathcal{L}_{\mathbb{B}}$-sentence".
- In this context, by the "extended language" we mean the language:

$$\mathcal{L}_{\mathbb{B}}(\mathrm{T}) := \mathcal{L}_{\mathbb{B}} \cup \{\mathrm{T}(x)\}$$

.

Gauging the logical distance between P[$\mathbb{B}$] and $\mathbb{B}$

- Tarski's undefinability of truth theorem and Gödel's second incompleteness theorem suggest that there is a significant logical gap between P[$\mathbb{B}$] and $\mathbb{B}$. We can measure this gap by answering the following questions.

- **Q1.** Is P[$\mathbb{B}$] *semantically conservative* over $\mathbb{B}$? In other words, does every model of $\mathbb{B}$ have an expansion to a model of P[$\mathbb{B}$]?

- **Q2.** Is P[$\mathbb{B}$] *(syntactically) conservative* over $\mathbb{B}$? In other words, if P[$\mathbb{B}$] $\vdash \varphi$, where φ is an $\mathcal{L}_{\mathbb{B}}$-sentence, then $\mathbb{B} \vdash \varphi$?

  The Tarski boundary demarcates the territory of truth theories that are conservative over $\mathbb{B}$.

  Note that semantic conservativity implies conservativity.

- **Q3.** Suppose P[$\mathbb{B}$] is conservative over $\mathbb{B}$. Among the canonical computational classes of functions $\mathcal{F}$, what is the *optimal complexity class* $\mathcal{F}$ that contains a function f such that for all proofs π and all $\mathcal{L}_{\mathbb{B}}$-sentences φ, we have: $\mathrm{P}[\mathbb{B}] \vdash_{\pi} \varphi \Longrightarrow \mathbb{B} \vdash_{f(\pi)} \varphi$.

- **Q4.** Is P[$\mathbb{B}$] *interpretable* in $\mathbb{B}$?

Conservativity of $\mathrm{CT}^-[\mathrm{PA}]$

- **Theorem.** $\mathrm{CT}^-[\mathrm{PA}]$ *is a conservative extension of* PA.

1. Krajewski, Kotlarski, and Lachlan [KKL] (1981) showed that every countable recursively saturated model of PA can be expanded to a model of $\mathrm{CS}^-[\mathrm{PA}]$ (where PA is treated as relational theory). They used a method that has come to be known as $\mathcal{M}$-logic (an exotic hybrid of proof theory and model theory). By elementary model theory this shows that CS^- is conservative over PA.
2. The $\mathcal{M}$-logic was extended by Kaye [Ka] (1991) and Engström [Eng] (2002) to languages with function symbols; the former in the context of CS^-, and the latter in the context of CT^-.
3. Visser and I [EV-1,2] (2012,2014) used the model-theoretic argument outlined in this lecture to show that $\mathrm{CT}^-[\mathrm{B}]$ is conservative over B for every base theory B (if B is formulated in a relational language.) We also showed if τ is a "scheme template" such that B proves every instance of τ, then: $\mathrm{CT}^-[\mathrm{B}] + \text{"every instance of } \tau \text{ is true"}$ is conservative over B.
4. The above model-theoretic method was further extended by Cieśliński [C-1] (2017) to obtain the conservativity of $\mathrm{CT}^-[\mathrm{PA}]$ over PA when PA is formulated in the usual functional language.
5. Leigh [L] (2015), using a newly minted cut-elimination argument to show that $\mathrm{CT}^-[\mathrm{B}]$ is conservative over B, where B is an extension of $\mathrm{I}\text{-}\Delta_0 + \mathrm{Exp}$ (in the same language).
6. Cieśliński [C-3] (2021) used classical techniques of proof theory and model theory to establish the conservativity of $\mathrm{CT}^-[\mathrm{PA}]$.

Conservative extensions of $\mathrm{CT}^-[\mathrm{PA}]$

- **Theorem.** *The following can be conservatively added to* $\mathrm{CT}^-[\mathrm{PA}]$

1. Int-Ind (internal induction) is the single sentence in the language $\mathcal{L}_{\mathrm{PA}}(\mathrm{T})$ that asserts that every instance of the induction scheme (of PA) is true, i.e., $\forall\varphi(x) \in \mathrm{Form}^1_{\mathrm{PA}}\ \mathrm{T}(\mathrm{Ind}_\varphi)$, where Ind_φ is the induction axiom for φ. In the presence of $\mathrm{CT}^-[\mathrm{PA}]$, Int-Ind is equivalent to the sentence asserting that all the axioms of the usual axiomatization of PA are true.
2. The collection of sentences $\forall x(\mathrm{True}_n(x) \to \mathrm{T}(x))$, where $n \in \omega$. Here $\mathrm{True}_n(x)$ is the definable truth predicate for Σ_n-sentences.
3. (DC-in) $\forall k \left(\exists i < k\ \mathrm{T}(\varphi_i) \longrightarrow \mathrm{T}(\bigvee_{i<k} \varphi_i) \right)$.
(Cieśliński, Łełyk, and Wcisło 2023).
4. Coll(T), where Coll(T) is the scheme of collection in the extended language (Wcisło 2024).

Interpretability of $\mathsf{CT}^-[\mathsf{PA}]$ in PA

- **Theorem.** (Visser + E. 2012) $\mathsf{CT}^-[\mathsf{PA}]$ *is interpretable in* PA.
- In contrast to the above theorem, Pudlák's interpretability theoretic generalization of the second incompleteness theorem (1985) can be used to show:
  (1) $\mathsf{CT}^-[\mathbb{B}]$ cannot be interpreted in $\mathbb{B}$ for a finiely axiomatized $\mathbb{B}$.
  (2) $\mathsf{CT}^-[\mathsf{PA}]$ + Int-Ind is not interpretable in PA.
- Visser (2019) has shown that no finitely axiomatizable base theory can $\mathbb{B}$ interpret $\mathsf{UTB}^-[\mathbb{B}]$.
- The question of whether a finitely axiomatized $\mathbb{B}$ can interpret $\mathsf{TB}^-[\mathbb{B}]$ can be weakened to $\mathsf{TB}^-[\mathbb{B}]$ here.

1. The interpretability of $\mathsf{CT}^-[\mathsf{PA}]$ in PA was first established in [EV-1], essentially the same proof is also presented in my recent paper [Ena-1]. Another proof was presented in my joint paper [EŁW] with Łełyk and Wcisło. Both proofs involve appropriate arithmetizations of the model-theoretic proof of conservativity of $\mathsf{CT}^-[\mathsf{PA}]$ over PA.
2. The interpretability of $\mathsf{CT}^-[\mathsf{PA}]$ in PA can also be derived from Leigh's proof-theoretic demonstration [L] of conservativity of $\mathsf{CT}^-[\mathsf{PA}]$ over PA.

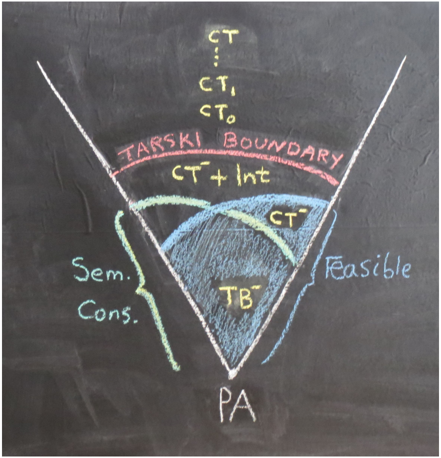
Feasible reducibility and hardness

- **Theorem.** (Łełyk + Wcisło + E. 2020). $\mathsf{CT}^-[\mathsf{PA}]$ *is feasibly reducible to* PA, *i.e., there is a polynomial-time computable function* f *with the property that for all proofs* π *and all* $\mathcal{L}_{\mathsf{PA}}$*-sentences* φ*, we have*:

$$\mathsf{CT}^-[\mathsf{PA}] \vdash_{\pi} \varphi \Longrightarrow \mathsf{PA} \vdash_{f(\pi)} \varphi.$$

- **Theorem.** (Łełyk and E. 2023) *The collection of* $\mathcal{L}_{\mathsf{PA}}(\mathsf{T})$*-sentences* θ *such that* $\mathsf{CT}^-[\mathsf{PA}] + \theta$ *is conservative over* PA *is complete* Π^0_2 (*and in particular, it is not recursively/computable enumerable*).

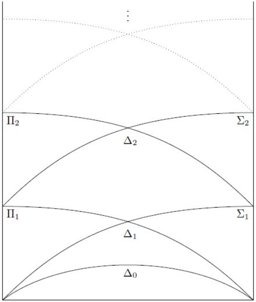
Tarski Boundary



Lecture 1 + Lecture 2 + Lecture 3

The Σ_n-hierarchy of formulas in arithmetic and set theory

- $\Sigma_0 = \Pi_0 = \Delta_0$ is the collection of $\mathcal{L}_{\mathrm{PA}}$-formulas all of whose quantifiers are of the form $\exists x < t\ \varphi$ or $\forall x < t\ \varphi$ (where t is a term).
- Σ_{n+1} consists of $\mathcal{L}_{\mathrm{PA}}$-formulas of the form $\exists x_0 \cdots \exists x_k\ \varphi$, where $\varphi \in \Pi_n$;
- Π_{n+1} consists of $\mathcal{L}_{\mathrm{PA}}$-formulas of the form $\forall x_0 \cdots \forall x_k\ \varphi$, where $\varphi \in \Sigma_n$.
- One can similarly define the (Levy) hierarchy of $\mathcal{L}_{\mathrm{ZF}}$-formulas, using $\in$ instead of $<$ for bounding quantifiers.



Connection with satisfaction classes

- **Theorem**. *For each $n \geq 1$ there is there is a Σ_n-formula* Sat_n *such that* PA *proves* $\mathrm{CS}^- \restriction \mathrm{F}$ *with* $\mathrm{S}(x, y)$ *replaced with* $\mathrm{Sat}_n(x, y)$ *and* $\mathrm{F}(x)$ *with "x is a Σ_n-formula".*

- For $\mathcal{M} \models \mathrm{PA}$, and $S \subseteq M^2$, we say that S is an F-satisfaction class over $\mathcal{M}$ if $(\mathcal{M}, S, F) \models \mathrm{CS}^- \restriction \mathrm{F}$.
  S is a full satisfaction class over $\mathcal{M}$ if $(\mathcal{M}, S) \models \mathrm{CS}^-$.

- **Corollary** *If $n \in \omega$ and $\mathcal{M} \models$* PA, *then there is an an F_n-satisfaction class over $\mathcal{M}$, where $F_n(x)$ iff $\mathcal{M} \models x \in \Sigma_n$.*

- MORAL: Despite Tarski's undefinability of truth theorem, arithmetical truth can be definably approximated.

- The above theorem can be used, together with the cut-elimination theorem for FOL, to establish the following twist to Gödel's second incompleteness theorem.

- **Theorem.** (Mostowski) PA *is a reflexive theory, i.e.,* PA *proves the consistency of each of its finite subtheories.*

The Basic Construction (1)

- We now proceed to present a proof-outline of a theorem that shows that $\mathsf{CS}^{-}[\mathsf{PA}]$ is conservative over PA.
- **Lemma.** *Let* $\mathcal{N}_0 \models \mathsf{PA}$, $F_1 := \mathrm{Form}^{\mathcal{N}_0}$, $F_0 \subseteq F_1$, *and* let S_0 *be an* F_0*-satisfaction class.* **Then** *there are* :
  1. $\mathcal{N}_1 \succ \mathcal{N}_0$, and
  2. *an* F_1*-satisfaction class* $S_1 \supseteq S_0$ over $\mathcal{N}_1$ *such that:*
     If $c \in F_0$, $\alpha \in N_0$, *and* $(c, \alpha) \in S_1$, **then** $(c, \alpha) \in S_0$
- **Proof:** Enrich $\mathcal{L}_{\mathsf{PA}}$ with constant symbols for each member of N_0, and new unary predicates U_c for each $c \in \mathrm{Form}^{\mathcal{N}_0}$. We will next formulate certain statements in the enriched language.

The Basic Construction (2)

- If $R \in \mathcal{L}_{\mathrm{PA}}$ and $\mathcal{N}_0 \models c = \ulcorner R(x_0, \cdots, x_{n-1}) \urcorner$, then

$$\theta_c := \forall \alpha \, (\mathrm{U}_c(\alpha) \leftrightarrow \mathrm{Asn}(\alpha, c) \wedge R(\alpha(x_0), \cdots, \alpha(x_{n-1}))).$$

- If $\mathcal{N}_0 \models c = \ulcorner \neg d \urcorner$, then $\theta_c := \forall \alpha \, (\mathrm{U}_c(\alpha) \leftrightarrow \mathrm{Asn}(\alpha, c) \wedge \neg \mathrm{U}_d(\alpha))$.

$$\theta_c := \forall \alpha \, (\mathrm{U}_c(\alpha) \leftrightarrow \mathrm{Asn}(\alpha, c) \wedge \neg \mathrm{U}_d(\alpha)).$$

- If $\mathcal{N}_0 \models c = \ulcorner d_1 \vee d_2 \urcorner$, then

$$\theta_c := \forall \alpha \, (\mathrm{U}_c(\alpha) \leftrightarrow \mathrm{Asn}(\alpha, c) \wedge (\mathrm{U}_{d_1}(\alpha \restriction \mathrm{FV}(d_1)) \vee \mathrm{U}_{d_2}(\alpha \restriction \mathrm{FV}(d_2)))).$$

- If $\mathcal{N}_0 \models c = \ulcorner \exists v \; b \urcorner$, then

$$\theta_c := \forall \alpha \, (\mathrm{U}_c(\alpha) \leftrightarrow \exists z \; \mathrm{U}_b(\alpha_z^v)).$$

The Basic Construction (3)

- Let $\Theta := \{\theta_c : c \in F_1\}$, and $\Gamma := \Gamma^+ \cup \Gamma^-$, where

  $\Gamma^+ := \{\mathsf{U}_c(\alpha) : c \in F_0 \text{ and } (c, \alpha) \in S_0\}$,

  $\Gamma^- := \{\neg\mathsf{U}_c(\alpha) : c \in F_0 \text{ and } (c, \alpha) \notin S_0\}$.

- Let $\mathrm{Th}^+(\mathcal{N}_0) := \mathrm{Th}(\mathcal{N}_0, a)_{a \in N_0} \cup \Theta \cup \Gamma$.

- Then we use a compactness argument to show that the desired $(\mathcal{N}, S)$ exists by verifying that each finite subset of $\mathrm{Th}^+(\mathcal{N}_0)$ is interpretable in $(\mathcal{N}_0, S_0)$.

The Basic Theorem

- **Theorem.** *Let $\mathcal{M}_0 \models \mathsf{PA}$. There is $\mathcal{M} \succ \mathcal{M}_0$ such that for some relation S over $\mathcal{M}$, $(\mathcal{M}, S) \models \mathsf{CS}^-$. Indeed S can be required to extend any prescribed partial satisfaction class over $\mathcal{M}$.*

- **Proof:** Suppose S_0 is a prescribed F_0-satisfaction class over $\mathcal{M}$. Then by the Basic Lemma we can build $\langle \mathcal{M}_i : i \in \omega \rangle$ and $\langle S_i : i \in \omega \rangle$ that satisfy the following three properties:

  1. $\mathcal{M}_i \prec \mathcal{M}_{i+1}$;
  2. S_{i+1} is an F_{i+1}-satisfaction class on $\mathcal{M}_{i+1}$ with $F_{i+1} := \mathsf{Form}^{\mathcal{M}_i}$;
  3. $S_i = S_{i+1} \cap \{(c, \alpha) : c \in F_i,\ \mathcal{M}_i \models \mathsf{Asn}(\alpha, c)\}$.

  Then, let $\mathcal{M} := \bigcup\limits_{i \in \omega} \mathcal{M}_i$, and $S := \bigcup\limits_{i \in \omega} S_i$.

- **Remark.** The above model-theoretic methodology can be further elaborated so as to allow the construction of models of $\mathsf{CT}^-[\mathsf{PA}]$ (via extensional satisfaction classes).

References for Lecture 1

C-1 C. Cieśliński, *On Some Problems with Truth and Satisfaction*, in **Philosophical Approaches to the Foundations of Logic and Mathematics**, Brill, 2021, pp. 175–192.

C-2 C. Cieśliński, **The Epistemic Lightness of Truth**. **Deflationism and its Logic**, Cambridge University Press, Cambridge, 2017.

C-3 C. Cieśliński, *Interpreting the compositional truth predicate in models of arithmetic*, vol. 60, pp. 749–770 (2021).

Ena A. Enayat, *Satisfaction classes with approximate disjunctive correctness*, **Review of Symbolic Logic**, vol. 18, pp. 545-562 (2025).

EŁ A. Enayat and M. Łełyk, *Axiomatizations of Peano Arithmetic: a truth-theoretic view*, **Journal of Symbolic Logic**, vol. 88 (2023), pp. 1526-1555.

EŁW A. Enayat, M. Łełyk, and B. Wcisło, *Truth and feasible reducibility*, **Journal of Symbolic Logic**, vol. 85 (2020), pp. 367–421.

EV-1 A. Enayat and A. Visser, *Full satisfaction classes in a general setting*, privately circulated manuscript (2012).

EV-2 A. Enayat and A. Visser, *New constructions of full satisfaction classes*, in **Unifying the Philosophy of Truth** (edited by D. Achourioti et al.), J. New York: Springer, pp. 321–325.

Eng F. Engström, *Satisfaction classes in nonstandard models of first-order arithmetic*, **arXiv:math/0209408** (2002).

H V. Halbach, **Axiomatic Theories of Truth,** 2^{nd}ed., Cambridge University Press, 2015.

Ka R. Kaye, R. Kaye, **Models of Peano Arithmetic**, Oxford Logic Guides, Oxford University Press, Oxford, 1991.

KKL H. Kotlarski, S. Krajewski, and A. H. Lachlan, *Construction of satisfaction classes for nonstandard models*, **Canadian Mathematical Bulletin,** vol. 24 (1981), pp. 283–293.

L G. Leigh, *Conservativity for theories of compositional truth via cut elimination*. **Journal of Symbolic Logic**, vol. 80 (2015), 845–865.

P P. Pudlák, *Cuts, consistency statements and interpretations*, **Journal of Symolic Logic**, vol. 50 (1985), pp. 423-441.

V A. Visser, *Enayat theories* **J. Appl. Logics** vol. 8 (2021), pp. 2073–2117.

W B. Wcisło, *Truth and Collection*, **arXiv:2403.19367** (2024).

END OF LECTURE ONE

Truth and consistency

- Recall that $\mathsf{CT}^{-}[\mathsf{PA}]$ is conservative over PA, so $\mathsf{CT}^{-}[\mathsf{PA}] \nvdash \mathrm{Con}(\mathsf{PA})$.
- On the other hand, one "should be able to" prove Con(PA) using the "argument" that:
  1. All the axioms of PA are true.
  2. T is closed under provability.
  3. 0=1 is not true.
  4. ERGO: 0=1 is not provable in PA.
- In the above, (3) is a theorem of $\mathsf{CT}^{-}[\mathsf{PA}]$, but neither (1) nor (2) is provable in $\mathsf{CT}^{-}[\mathsf{PA}]$.
- However both (1) and (2) are readily provable in CT[PA] $:= \mathsf{CT}^{-}[\mathsf{PA}] + \mathrm{Ind}(\mathsf{T})$.

The fascinating theory CT_0

- $\mathrm{GRef}_{\mathrm{PA}}$ (Golbal Reflection over PA) denotes sentence :

$$\forall \varphi \in \mathrm{Sent}_{\mathrm{PA}} \; [\mathrm{Prov}_{\mathrm{PA}}(\varphi) \rightarrow \mathrm{T}(\varphi)].$$

- **Theorem.** $\mathrm{CT}[\mathrm{PA}] \vdash \mathrm{GRef}_{\mathrm{PA}}$.
- **Corollary.** $\mathrm{CT}[\mathrm{PA}] \vdash \mathrm{Con}_{\mathrm{PA}},\; \mathrm{Con}_{\mathrm{PA}+\mathrm{Con}_{\mathrm{PA}}}$, etc.
- Let $\mathrm{CT}_n[\mathrm{PA}] := \mathrm{CT}^-[\mathrm{PA}] + \Sigma_n\text{-}\mathrm{Ind}(\mathrm{T})$.
- **Theorem.** $\mathrm{CT}_1[\mathrm{PA}] \vdash \mathrm{GRef}_{\mathrm{PA}}$.
- What about $\mathrm{CT}_0[\mathrm{PA}]$?

The Many Faces Theorem

- **Many Faces Theorem of** $\mathsf{CT}_0[\mathsf{PA}]$ (1986 $\rightarrow$ 2023).
  (Kotlarski, Cieśliński, Wcisło, Łełyk, Pakhomov, and E.)

  The following axiomatize the same theory over $\mathsf{CT}^-[\mathrm{I}\Delta_0 + \mathsf{Exp}]$.

  1. Δ_0-Ind(T).
  2. $\mathsf{GRef}_{\mathsf{PA}}$.
  3. *The sentence asserting that* T *is closed under first order proofs.*
  4. *The sentence asserting that* T *contains all* $\mathcal{L}_{\mathsf{PA}}$*-instances of theorems of first order logic.*
  5. *The sentence asserting that* T *is closed under proofs of propositional logic.*
  6. (DC) $\forall k \left(\mathsf{T}(\bigvee_{i<k} \varphi_i) \longleftrightarrow \exists i < k \; \mathsf{T}(\varphi_i) \right)$.
  7. (DC-out) $\forall k \left(\mathsf{T}(\bigvee_{i<k} \varphi_i) \longrightarrow \exists i < k \; \mathsf{T}(\varphi_i) \right)$.

Axiomatizations of ZF

- ZF (Zermelo-Fraeknel) set theory is usually axiomatized by adding the Replacement scheme to finitely many axioms: Extensionality, Emptyset, Foundation (Regularity), Pair, Union, Powerset, and infinity.
- In the presence of the other axioms, the replacement scheme can be interchanged with the combination of the Separation scheme and the Collection scheme.
- The Separation scheme consists of the sentences of following form:

$$\forall v \ \forall b \ \exists a \ \forall x \, (x \in a \ \leftrightarrow \ \varphi(x, v)) \, .$$

- The Collection scheme consists of the sentences of following form:

$$\forall v \ [(\forall x \in a \ \exists y \ \psi(x, y, v)) \rightarrow (\exists b \ \forall x \in a \ \exists y \in b \ \psi(x, y, v))].$$

- Let $\mathsf{ZF}^{-\infty} := \mathsf{ZF} \setminus \{\mathsf{Infinity}\} + \neg\mathsf{Infinity} + \mathsf{TC}$, where TC is the statement "every set has a transitive closure".
  1. The intended model of $\mathsf{ZF}^{-\infty}$ is $(V_\omega, \in)$, where:
  $$V_0 = \varnothing, \quad V_{n+1} = \mathcal{P}(V_n), \quad \text{and } V_\omega := \bigcup_{n\in\omega} V_n.$$
  2. $\mathrm{Ord} = \omega$ within $\mathsf{ZF}^{-\infty}$, thus the natural numbers form a proper class in this theory.
- **Theorem.** (Ackermann 1940, Mycielski 1964, Kaye-Wong 2007) PA *and* $\mathsf{ZF}^{-\infty}$ *are definitionally equivalent. In particular, they are bi-interpretable.*

$\mathrm{GB}^{\pm\infty}$

- GB is Gödel–Bernays theory of classes (also known as BG, VNB and NBG).
- Let $\mathrm{GB}^{-\infty} := \mathrm{GB} \setminus \{\mathrm{Infinity}\} + \neg\mathrm{Infinity} + \mathrm{TC}$.
- Models of $\mathrm{GB}^{\pm\infty}$ can be written in the two-sorted form $(\mathcal{M}, \mathfrak{X})$, where $\mathcal{M} \models \mathrm{ZF}^{\pm\infty}$, and $\mathfrak{X} \subseteq \mathcal{P}(M)$.
- $(\mathcal{M}, \mathfrak{X}) \models \mathrm{GB}^{\pm\infty}$ iff the following two conditions hold:
  1. If $X_1, \cdots, X_n \in \mathfrak{X}$, and Y is parametrically definable in $(\mathcal{M}, X_1, \cdots, X_n)$, then $Y \in \mathfrak{X}$.
  2. If $X_1, \cdots, X_n \in \mathfrak{X}$, then $(\mathcal{M}, X_1, \cdots, X_n) \models \mathrm{ZF}^{\pm\infty}(X_1, \cdots, X_n)$.

  In the above $\mathrm{ZF}^{\pm\infty}(X_1, \cdots, X_n)$ is the extension of $\mathrm{ZF}^{\pm\infty}$ in which the names for $X_1, \cdots, X_n$ can be used in the replacement scheme.

Useful facts about $\mathrm{GB}^{\pm\infty}$

1. **Theorem.** (Novak and Mostowski 1950) $\mathrm{GB}^{\pm\infty}$ *is conservative over* $\mathrm{ZF}^{\pm\infty}$.

   The original proof of Theorem 1 was model-theoretic. With refined arithmetization, it can implemented in the fragment $I\Sigma_1$ of PA (and therefore it is provable in PRA).

   Shoenfield (1954) used ε-calculus to estabish Theorem 1. His proof can be implemented in $\mathrm{I}\Delta_0 + \mathrm{Supexp}$.

2. **Theorem.** (Pudlák 1986, Solovay) $\mathrm{GB}^{\pm\infty}$ *has superexponential speed-up over* $\mathrm{ZF}^{\pm\infty}$.

3. **Theorem.** (Folklore) $\mathrm{ZF}^{\pm\infty}$ *is interpretable in* $\mathrm{GB}^{\pm\infty}$*, but not vice-versa.*

4. **Theorem.** $\mathrm{GB}^{-\infty}$ *and* ACA_0 *are definitionally equivalent. In particular, they are bi-interpretable.*

   The above theorem can be viewed as a corollary of the definitional equivalence of $\mathrm{ZF}^{-\infty}$ and PA.

The Mostowski truth predicate

- Mostowski (1950) came up with an elegant construction that shows another dramatic difference between $\mathrm{ZF}^{\pm\infty}$ and $\mathrm{GB}^{\pm\infty}$, namely, there is a formula $\mathsf{T}_{\mathsf{Most}}(x)$ – dubbed the *Mostowski truth predicate* here – such that for all sentences φ in the language of ZF-set theory, we have:

$$\mathrm{GB}^{\pm\infty} \vdash \varphi \leftrightarrow \mathsf{T}_{\mathsf{Most}}(\ulcorner\varphi\urcorner).$$

- The conservativity of $\mathrm{GB}^{\pm\infty}$ over $\mathrm{ZF}^{\pm\infty}$, combined with Gödel's second incompleteness theorem makes it clear that $\mathrm{Con}(\mathrm{ZF}^{\pm\infty})$ is unprovable in GB.
- Thus we are faced with a striking phenomenon: $\mathrm{GB}^{\pm\infty}$ possesses a truth-predicate for ZF, and yet the formal consistency of $\mathrm{ZF}^{\pm\infty}$ is unprovable in GB, which must be due to the lack of sufficient formal induction available in $\mathrm{GB}^{\pm\infty}$ in order to prove the statement "all theorems of $\mathrm{ZF}^{\pm\infty}$ are true".

Detour: $\mathsf{CT}^- \upharpoonright \mathsf{F}$ for set theory

- Within $\mathsf{ZF}^{\pm\infty}$, let $\mathcal{L}^+_{\mathsf{ZF}}$ be the result of augmenting $\mathcal{L}_{\mathsf{ZF}}$ with a constant symbol $\dot{x}$ for each x in the universe.
- Let $\mathcal{M} \models \mathsf{ZF}^{\pm\infty}$, and suppose $F \subseteq \mathsf{Form}^{\mathcal{M}}_{\mathsf{ZF}}$. $\mathsf{FSent}^{(\mathcal{M},F)}$ consists of $m \in M$ such that $(\mathcal{M}, F)$ satisfies:

m codes an $\mathcal{L}^+_{\mathsf{ZF}}$-sentence $\varphi(\dot{a}_1, ..., \dot{a}_k)$ for some $\varphi(x_1, ..., x_k) \in \mathsf{F}$.

- T is an F-truth class (over $\mathcal{M}$) if $(\mathcal{M}, F, T) \models \mathsf{CT}^- \upharpoonright \mathsf{F}$, where $\mathsf{CT}^- \upharpoonright \mathsf{F}$ is a finitely axiomatized theory that stipulates that T satisfies Tarski's compositional clauses at least for formulas in F (the axioms are given on the next page).

Detour concluded: Truth classes

- The axioms of $\mathsf{CT}^- \restriction \mathsf{F}$ are listed below.

  (1) $\forall x, y\ [\mathsf{F}(x) \to [\mathsf{Form}_{\mathsf{ZF}}(x) \wedge [y \lhd x \ \to \mathsf{F}(y)]]$.

  (2) $\forall x, y\, [\mathsf{T}(x) \to \mathsf{FSent}(x)]$.

  (3) $\forall x, y \left(\ulcorner \dot{x} = \dot{y} \urcorner \in \mathsf{T} \leftrightarrow x = y \right) \wedge \left(\ulcorner \dot{x} \in \dot{y} \urcorner \in \mathsf{T} \leftrightarrow x \in y \right)$.

  (4) $\forall \varphi, \psi \in \mathsf{FSent}\, [(\varphi = \neg\psi) \to (\varphi \in \mathsf{T} \leftrightarrow \psi \notin \mathsf{T})]$.

  (5) $\forall \varphi, \psi_1, \psi_2 \in \mathsf{FSent}$
  $[(\varphi = \psi_1 \vee \psi_2) \to (\varphi \in \mathsf{T} \leftrightarrow ((\psi_1 \in \mathsf{T}) \vee (\psi_1 \in \mathsf{T})))]$.

  (6) $\forall \varphi \in \mathsf{FSent},\ \forall \psi(v) \in \mathsf{F}^1$
  $\left[(\varphi = \exists v\ \psi(v)) \to \left(\varphi \in \mathsf{T} \leftrightarrow \exists x\ \psi(\dot{x}/v) \in \mathsf{T} \right) \right]$.

- $\mathsf{CT}^- := \mathsf{CT}^- \restriction \mathsf{F} + (\mathsf{F} = \mathsf{Form}_{\mathsf{ZF}})$.
  Thus in this context, CT^- can be formulated in the language $\mathcal{L}_{\mathsf{ZF}}(\mathsf{T})$ (with no mention of F).

Back to the Mostowski Truth predicate

- A cut of a nonstandard ω-model $\mathcal{M}$ of $\mathsf{ZF}^{\pm\infty}$ is an initial segment of $\omega^{\mathcal{M}}$ that is closed under immediate successors.
- Let $\mathrm{depth}(\varphi)$ be the length of the longest path in the parsing tree of φ (also known as the formation/syntactic tree).
- Let Depth_k be the collection of $\mathcal{L}_{\mathsf{set}}$-formulae φ with $\mathrm{depth}(\varphi) \leq k$. Thus, $\mathrm{Depth}_0 = \varnothing$, and Depth_1 consists of atomic formulae.
- Within $\mathsf{GB}^{\pm\infty}$, the Mostowski cut, denoted $\mathsf{C}_{\mathsf{Most}}$, consists of $k \in \omega$ such that there is a class T with the property that T is a Depth_k-truth class over the structure $(\mathbf{V}, \in)$.
- **Lemma.** *Provably in* $\mathsf{GB}^{\pm\infty}$, $\mathsf{C}_{\mathsf{Most}}$ *is a cut of* ω.
- **Theorem.** *If* $\mathcal{M}$ *is an* ω*-nonstandard model of* $\mathsf{ZF}^{\pm\infty}$*, then* $\mathsf{C}_{\mathsf{Most}}$*-as-calculated-in-*$(\mathcal{M}, \mathrm{Def}(\mathcal{M}))$ *coincides with the standard cut of* $\mathcal{M}$*.*
- **Corollary.** *The statement* $\mathsf{C}_{\mathsf{Most}} = \omega$ *is unprovable in* $\mathsf{GB}^{\pm\infty}$*. Therefore, there is an instance of* Σ^1_1-Ind *that is not provable in* $\mathsf{GB}^{\pm\infty}$.

The Mostowski truth predicate concluded

- Within $\mathrm{GB}^{\pm\infty}$, we define:
  (a) $\mathrm{Depth}_{\mathsf{C}_{\mathsf{Most}}}$ consists of $\mathcal{L}_{\mathsf{ZF}}$-formulae φ such that $\mathrm{depth}(\varphi) \in \mathsf{C}_{\mathsf{Most}}$.
  (b) The Mostowski truth predicate, denoted $T_{\mathsf{Most}}(x)$ expresses:

$$x \text{ is (the code of) an } \mathcal{L}^{+}_{\mathsf{ZF}}\text{-formula } \varphi(\dot{a}_1, ..., \dot{a}_k) \in \mathrm{Depth}_k, \text{ and}$$

$$\exists\, p \geq k\ \exists T\ \varphi(\dot{a}_1, ..., \dot{a}_k) \in T \text{ for some } \mathrm{Depth}_p\text{-truth class } T.$$

- **Theorem.** (Mostowski 1950).

$$\mathrm{GB}^{\pm\infty} \quad \vdash \quad \mathsf{T}_{\mathsf{Most}}(x) \text{ is a } \mathrm{Depth}_{\mathsf{C}_{\mathsf{Most}}}\text{- truth predicate.}$$

Consequently, $\mathrm{GB}^{\pm\infty} \vdash \varphi \leftrightarrow \mathsf{T}_{\mathsf{Most}}(\ulcorner\varphi\urcorner)$, for all $\mathcal{L}_{\mathsf{ZF}}$-sentences φ.

- **Theorem.** (Solovay 1976). *Provably in* $\mathrm{GB}^{\pm\infty}$, $\mathrm{Con}(\mathrm{ZF}^{\pm\infty})$ *holds in* $\mathsf{C}_{\mathsf{Most}}$.

Connections with extensions of ACA_0

- $\mathsf{ACA}_0' := \mathsf{ACA}_0 + \forall k\, \forall X\, \exists Y\ Y = X^{(k)}$,
  where $Y = X^{(k)}$ expresses: Y is the k-th Turing jump of X.
- $\mathsf{ACA}_0^* := \mathsf{ACA}_0 + \forall k\, \exists Y\ Y = 0^{(k)}$.
- ACA_0' and ACA_0^* have the same purely arithmetical consequences.
- $\mathsf{GB}_*^{\pm\infty} := \mathsf{GB}^{\pm\infty} + \mathsf{C}_{\mathsf{Most}} = \omega$.
- **Theorem.** $\mathsf{GB}_*^{-\infty}$ *is definitionally equivalent with* ACA_0^*.
- **Theorem.** *The following is provable in* $\mathsf{GB}_*^{\pm\infty}$

$$\forall \varphi \in \mathsf{Sent}_{\mathsf{ZF}}\ [\mathsf{Prov}_{\mathsf{ZF}^{\pm\infty}}(\varphi) \to \mathsf{T}_{\mathsf{Most}}(\varphi)].$$

- **Corollary.** $\mathsf{GB}_*^{\pm\infty} \vdash \mathsf{Con}_{\mathsf{ZF}^{\pm\infty}}$.

Connections with $\mathsf{REF}^{\omega}(\mathsf{PA})$.

- Given a recursively axiomatized theory U extending $\mathrm{I}\Delta_0 + \mathsf{Exp}$, the *uniform reflection scheme over* U, denoted $\mathsf{REF}(U)$, is defined via:

$$\mathsf{REF}(U) := \{\forall x(\mathsf{Prov}_U(\ulcorner \varphi(\dot{x}) \urcorner) \to \varphi(x)) : \varphi(x) \in \mathsf{Form}_U\}.$$

- The sequence of schemes $\mathsf{REF}^{\alpha}(U)$, where α is a recursive ordinal, is defined as follows:

$\mathsf{REF}^{0}(U) = U$;

$\mathsf{REF}^{\alpha+1}(U) = \mathsf{REF}(\mathsf{REF}^{\alpha}(U))$;

$\mathsf{REF}^{\gamma}(U) = \bigcup_{\alpha<\gamma} \mathsf{REF}^{\alpha}(U)$.

Back to $\mathsf{CT}_0[\mathsf{PA}]$!!!

- **Theorem A.** (McAloon 1985) *The $\mathcal{L}_{\mathsf{PA}}$-consequences of* ACA_0' *is axiomatized by* $\mathsf{REF}^{\omega}(\mathsf{PA})$.
- **Theorem B.** (Kotlarski 1986 + Wcisło 2017) *The $\mathcal{L}_{\mathsf{PA}}$-consequences of* $\mathsf{CT}_0[\mathsf{PA}]$ *is axiomatized by* $\mathsf{REF}^{\omega}(\mathsf{PA})$.
- **Theorem C.** ACA_0', ACA^*, *and* $\mathsf{CT}_0[\mathsf{PA}]$ *are pairwise mutually ω-interpretable; Consequently, they have the same $\mathcal{L}_{\mathsf{PA}}$-consequences.*
- Theorem C allows us to infer Theorems A from B from each other.

Proof outline of Theorem C

Since $\mathsf{ACA}_0^{'} \vdash \mathsf{ACA}_0^{*}$, it suffices to show:

$$\mathsf{ACA}_0^{*} \rhd_\omega \mathsf{CT}_0[\mathsf{PA}] \rhd_\omega \mathsf{ACA}_0^{'}.$$

To verify that $\mathsf{ACA}_0^{*} \rhd_\omega \mathsf{CT}_0[\mathsf{PA}]$, within ACA_0^{*} interpret the truth predicate T of $\mathsf{CT}_0[\mathsf{PA}]$ as $\mathsf{T}_{\mathsf{Most}}$.

To give an idea of the proof of $\mathsf{CT}_0[\mathsf{PA}] \rhd_\omega \mathsf{ACA}_0^{'}$, we describe the interpretation model-theoretically:
Given a model $(\mathcal{M}, T) \models \mathsf{CT}_0[\mathsf{PA}]$, let $\mathfrak{X}_{\mathrm{Def}_T(\mathcal{M})}$ be the collection of parametrically definable subsets of $\mathcal{M}$ from the point of view of T, i.e.,

$$\mathfrak{X}_{\mathrm{Def}_T(\mathcal{M})} := \{X \subseteq M : X = \varphi^T \text{ for some unary } \varphi(x) \in \mathsf{Form}^{\mathcal{M}}_{\mathsf{PA}}\},$$

where $\varphi^T := \left\{ m \in M : \varphi(\dot{m}) \in T \right\}$. One then checks that:

$$(\mathcal{M}, \mathfrak{X}_{\mathrm{Def}_T(\mathcal{M})}) \models \mathsf{ACA}_0'.$$

Historical notes for Lecture 2

- The story of the Many Faces Theorem of CT_0 started with Kotlarski [K] (1986), when he unveiled the connection with $GREF_{PA}$ by using an ingenious translation to show that $CT^-[PA] + GREF_{PA} \vdash CT_0[PA]$.
- Kotlarski's paper [K] also includes a proof outline for showing that that, conversely, $CT_0[PA] \vdash GREF_{PA}$, but a serious gap was found in his proof outline around 2012 by Heck and Visser.
- Wcisło [WŁ] (2017) used a different line of argument to show that $CT_0[PA]$ and Ref_{PA} have the same arithmetic consequences.
- Eventually Łełyk [Ł] (2018, 2023) confirmed Kotlarski's hunch by appropriate bootstrapping to verify that show that indeed $CT_0[PA] \vdash GREF_{PA}$.
- Kotlarski's work was revisited by Cieśliński [C] (2012), who showed that seemingly weak principles such as the principle "T-is closed under propositional reasoning" is equivalent to $\Delta_0(T)$-Ind over $CT^-[PA]$. His work, in turn, was refined by Pakhomov and E. [EP] (2019) in the demonstration of the equivalence of DC (Disjunctive Correctness) with $\Delta_0(T)$-Ind over $CT^-[PA]$. This result, in turn was further refined in [CŁW] (2023) to show that DC_{out} is equivalent to $\Delta_0(T)$-Ind over $CT^-[PA]$.
- A proof-theoretic demonstration of Theorem A were presented by Afshari and Rathjen [AR].
- Łełyk's paper [Ł] includes a model-theoretic proof of $CT_0[PA] \vdash REF^{\omega}(PA)$. A proof-theoretic demonstration of Theorem B was presented by Beklemishev and Pakhomov [BP].
- A proof outline of Theorem C is presented in [E], which revisits the various aspects of the Mostowski truth predicate and makes the connection between ACA_0' and $CT_0[PA]$ explicit.

References for Lecture 2

AR B. Afshari and M. Rathjen, *Ordinal Analysis and the Infinite Ramsey Theorem*, in **How the World Computes** (edited by B. Cooper et al.), CiE 2012, Lecture Notes in Computer Science, vol 7318, Springer, Berlin, Heidelberg, 2012, pp. 1-10.

BP L. D. Beklemishev and F. N. Pakhomov, *Reflection algebras and conservation results for theories of iterated truth,* **Annals of Pure and Applied Logic,** vol. 173(5), 103093 (2022), 41 pp.

C C. Cieśliński, *Truth, Conservativeness and Provability*, **Mind**, vol. 119 (2010), pp. 409-422.

CŁW C. Cieśliński, M. Łełyk, and B. Wcisło, *The two halves of disjunctive correctness*, **Jounal of Mathematical Logic**, vol. 23(2), Article no. 2250026, (2023), 28 pp.

E A. Enayat, *The Mostowski Bridge*, **arXiv:2505.23998** (2025).

EP A. Enayat and F. Pakhomov, *Truth, disjunction, and induction*, **Archive for Mathematical Logic**, vol. 58 (2019) , pp. 753–766.

Ma K. McAloon, *Paris-Harrington incompleteness and progressions of theories*, in **Proceedings of Symposia in Pure Mathematics,** vol. 42, pp. 447–460, 1985.

K H. Kotlarski, *Bounded induction and satisfaction classes*, **Mathematical Logic Quarterly.** vol. 32 (1986), pp. 531–544.

Ł M. Łełyk, *Model theory and proof theory of the global reflection principle*, **Journal of Symbolic Logic,** vol. 88, 738–779,(2023).

Mo Mostowski-impredicative A. Mostowski, *Some impredicatve definitions in the axiomatic set-theory,* **Fundamenta Mathematicae,** vol. 37 (1950), pp. 111-124.

S J. Shoenfield, *A Relative Consistency Proof*, **Journal of Symbolic Logic** vol. 19 (1954), pp. 21-28.

WŁ B. Wcisło and M. Łełyk, *Notes on bounded induction for the compositional truth predicate*, **The Review of Symbolic Logic**, vol. 10 (2017), pp. 455–480.

END OF LECTURE TWO

(Full) Compositional Truth over ZF

- Within ZF, let $\mathcal{L}^+_{\mathrm{ZF}}$ be the result of augmenting $\mathcal{L}_{\mathrm{ZF}}$ with a constant symbol $\dot{x}$ for each object x in the universe; e.g., let $\dot{x} := \langle x, 0\rangle$. We use $\mathrm{Sent}^+_{\mathrm{ZF}}$ to denote the class of $\mathcal{L}^+_{\mathrm{ZF}}$-sentences
- In our context (set theory) CT^- is the theory formulated in $\mathcal{L}_{\mathrm{ZF}}(\mathrm{T})$ whose axioms are as follows.
  1. $\forall x\ [\mathrm{T}(x) \rightarrow \mathrm{Sent}^+_{\mathrm{ZF}}(x)]$.
  2. $\forall x\ \forall y\ \left[\mathrm{T}(\dot{x} = \dot{y}) \leftrightarrow (x = y)\right]$.
  3. $\forall x\ \forall y\ \left[\mathrm{T}(\dot{x} \in \dot{y}) \leftrightarrow (x \in y)\right]$.
  4. $\forall \varphi \in \mathrm{Sent}^+_{\mathrm{ZF}}\ \left[\mathrm{T}(\neg\varphi) \leftrightarrow \neg\mathrm{T}(\varphi)\right]$.
  5. $\forall \varphi, \theta \in \mathrm{Sent}^+_{\mathrm{ZF}}\ \left[\mathrm{T}(\varphi \vee \theta) \leftrightarrow \mathrm{T}(\varphi) \vee \mathrm{T}(\theta)\right]$.
  6. $\forall \psi \in \mathrm{Sent}^+_{\mathrm{ZF}}\ [\psi = \exists v \varphi(v)] \rightarrow \left[\mathrm{T}(\exists v\ \varphi(v)) \leftrightarrow \exists x \mathrm{T}(\varphi(\dot{x})\right]$.
- $\mathrm{CT} := \mathrm{CT}^- + \mathrm{Repl}(\mathrm{T})$, where Repl(T) is the replacement scheme in the extended language. Repl(T) can be replaced with $\mathrm{Sep}(\mathrm{T}) + \mathrm{Coll}(\mathrm{T})$, where Sep(T) is the separation scheme for the extended language, and Coll(T) is the collection scheme for the extended language.

Lecture 1 + Lecture 2 + Lecture 3

Examples

- Recall that $\mathsf{CT}^-[\mathsf{ZF}] := \mathsf{ZF} + \mathsf{CT}^-$, and $\mathsf{CT}[\mathsf{ZF}] := \mathsf{CT}^-[\mathsf{ZF}] + \mathsf{Sep}(\mathsf{T}) + \mathsf{Coll}(\mathsf{T})$.

1. If $\mathcal{M} \models \mathsf{ZF}$ is ω-standard, then $(\mathcal{M}, \mathsf{Th}(\mathcal{M}^+)) \models \mathsf{CT}^-[\mathsf{ZF}] + \mathsf{Ind}(\mathsf{T})$.
2. If κ is strongly inaccessible , then $(V_\kappa, \in, \mathsf{Th}(V_\kappa, \in)^+) \models \mathsf{CT}[\mathsf{ZF}]$.
3. If $\mathcal{M}$ is either the first $(V_\alpha, \in) \models \mathsf{ZF}$, or $\mathcal{M}$ is the Shepherdson-Cohen minimal model of ZF. Then $(\mathcal{M}, \mathsf{Th}(\mathcal{M}^+) \models \mathsf{CT}^-[\mathsf{ZF}]$, but NOT $\mathsf{CT}[\mathsf{ZF}]$.
4. If $\mathcal{M} \models \mathsf{ZF}$ is ω-nonstandard, then $(\mathcal{M}, \mathsf{Th}(\mathcal{M}^+))$ is NOT a model of CT^-.
5. Generally speaking, the task of building a subset T of a given ω-nonstandard model $\mathcal{M}$ of ZF such that $(\mathcal{M}, T) \models \mathsf{CT}^-[\mathsf{ZF}]$ is highly nontrivial, and often impossible (e.g., if $\mathcal{M}$ is not recursively saturated).
6. In the above, and elsewhere, we use the convention introduced in Lecture 1 (page 5) of denoting the expansion $(\mathcal{M}, m)_{m \in M}$ of a structure $\mathcal{M}$ by $\mathcal{M}^+$. Thus $Th(\mathcal{M}^+)$ is the elementary diagram of $\mathcal{M}^+$.

Tarski Boundary (1)

- Recall that the Tarski Boundary separates conservative theories of truth (over a given base theory such as ZF) from nonconservative ones.
- **Theorem.** (Fujimoto 2012) *The following are provable in* CT[ZF]:
  1. *The sentence asserting*: there are arbitrarily large ordinals α such that as seen by T, $(V_\alpha, \in) \prec (V, \in)$.
  2. $\mathrm{GRef_{ZF}} := \forall \varphi \in \mathrm{Sent_{ZF}}\ [\mathrm{Prov_{ZF}}(\varphi) \to \mathrm{T}(\varphi)]$.
- **Corollary.** CT[ZF] proves that ZF has models of the form $(V_\alpha, \in)$.
- Let $\mathrm{CT}_n[\mathrm{ZF}] := \mathrm{CT}^-[\mathrm{ZF}] + \mathrm{Sep_n(T)} + \mathrm{Coll_n(T)}$, where $\mathrm{Sep_n(T)}$ and $\mathrm{Coll_n(T)}$ are the result (respectively) of limiting Sep(T) and Coll(T) to Σ_n-formulas.
- I have recently shown Fujimoto's proof can be refined so as to show that in the above theorem and corollary, CT[ZF] can be weakened to $\mathrm{CT}_0[\mathrm{ZF}]$.

Conservativity of $\mathsf{CT}^-[\mathsf{ZF}] + \mathsf{Sep}(\mathsf{T})$ (1)

- **Theorem.** (Essentially Krajewski 1975). $\mathsf{CT}^-[\mathsf{ZF}] + \mathsf{Sep}(\mathsf{T})$ *is conservative over* ZF.

  The proof of Krajewski's theorem can be combined with the Orey Compactness Theorem to show that $\mathsf{CT}^-[\mathsf{ZF}] + \mathsf{Sep}(\mathsf{T})$ is interpretable in ZF.

- Since Ind(T) (the full scheme of induction over natural numbers in the extended language) is provable in $\mathsf{CT}^-[\mathsf{ZF}] + \mathsf{Sep}(\mathsf{T})$, The above theorem implies:

- **Corollary.** $\mathsf{CT}^-[\mathsf{ZF}]$ + "T *is closed under first order proofs*" *is conservative over* ZF.

- The above result is in sharp contrast with the well-known fact (discussed in Lecture 2) that $\mathsf{CT}^-[\mathsf{PA}]$ + "T is closed under first order proofs" is not conservative over PA since it proves $\mathrm{Con}_{\mathrm{PA}}$

Proof outline of Krajewski's conservativity theorem

- The proof we will present below is a model-theoretic variant of Krajewski's original proof.
- Recall that the classical **Montague-Vaught reflection theorem** (1959) states that for all $\mathcal{L}_{\mathsf{ZF}}$ formula $\varphi(x_1, \cdots x_k)$, ZF can prove that there are arbitrarily large ordinals α such that $(V_\alpha, \in) \prec_\varphi (V, \in)$, i.e.,

$$\mathsf{ZF} \vdash \forall x_1 \in V_\alpha \cdots \forall x_k \in V_\alpha \left[\varphi^{V_\alpha}(x_1 \cdots \cdot x_k) \leftrightarrow \varphi(x_1 \cdots x_k)\right].$$

- It suffices to show that every recursively saturated model of ZF has an elementary submodel that can be expanded to $\mathsf{CT}^-[\mathsf{ZF}] + \mathsf{Sep(T)}$.
  Let $\mathcal{M}$ be a recursively saturated extension of $\mathcal{M}$.
  The reflection theorem can be used to show that there is some $\alpha \in \mathsf{Ord}^{\mathcal{M}}$ such that $\mathcal{M}_\alpha \prec \mathcal{M}$, where $\mathcal{M}_\alpha := (V_\alpha, \in)^{\mathcal{M}}$.
  Let T be $\mathsf{Th}(V_\alpha, \in)^+$-as-calculated-in-$\mathcal{M}$.
  Then $(\mathcal{M}_\alpha, T) \models \mathsf{CT}^-[\mathsf{ZF}] + \mathsf{Sep(T)}$.

Conservativity of $CT^-[ZF]$ + "the axioms of ZF are true"

- Int-Repl (internal replacement) is the axiom in the language $\mathcal{L}_{ZF}(T)$ that states that all instances of the replacement scheme of ZF are true.
- If ZF is axiomatized by adding finitely many axioms to the replacement scheme, then in the presence of $CT^-[ZF]$:

  Int-Repl is equivalent to the $\mathcal{L}_{ZF}(T)$-sentence that states that every axiom of ZF is true.
- **Theorem.** (Visser and E. 2012) $CT^-[ZF]+$ Int-Repl *is conservative over* ZF.
- Even though each of the theories $CT^-[ZF] + Sep(T)$, and $CT^-[ZF] +$ Int-Repl is conservative over ZF, their union proves Con_{ZF} and is thus nonconservative.
- In contrast with $CT^-[ZF] + Sep(T)$, $CT^-[ZF]$ + Int-Repl is not interpretable in ZF. This follows from Pudlák's interpretability-theoretic generalization of the second incompleteness theorem.

A new conservativity result

- In the remaining slides, I will present some recently obtained results (to appear).
- **Theorem.** *It is a theorem of* ZFC *that* $\mathrm{CT}^-[\mathrm{ZF}]$ +Int-Repl + Coll(T) *is conservative over* ZF.

  Proof-idea. An $\mathcal{L}_{\mathrm{ZF}}$-structure $\mathcal{M}$ is said to be $\aleph_1$-like if the universe M of $\mathcal{M}$ has cardinality $\aleph_1$, but for each $a \in M$, $\{x \in M : \mathcal{M} \models x \in a\}$ is countable. Using the fact that $\aleph_1$ is a regular cardinal (in ZFC), it is readily seen that if $\mathcal{M}$ is $\aleph_1$-like, then $\mathcal{M}$ satisfies Coll($\mathcal{L}$).

  By the completeness theorem of first order logic, it suffices to show that every countable model $\mathcal{M}_0 \models \mathrm{ZF}$ has an $\aleph_1$-like elementary extension $\mathcal{M}$ that has an expansion $(\mathcal{M}, T) \models \mathrm{CT}^-[\mathrm{ZF}]$ +Int-Repl + Coll(T).

  The main technical ideas of the proof are:
  (1) Keisler's theorem on elementary end extensions of models of ZF,
  (2) the conservativity of $\mathrm{CT}^-[\mathrm{ZF}]$ + Int-Coll over ZF, and
  (3) the "cutting-extending-sowing" trick used by Wcisło in his recent proof of conservativity of $\mathrm{CT}^-[\mathrm{PA}]$ + Coll over PA.
- **Question.** Can the meta-theory in the above theorem be reduced to PA?

Nonconservative theories weaker than $\mathsf{CT}_0[\mathsf{ZF}]$.

- We already saw that $\mathsf{CT}_0[\mathsf{ZF}]$ is on the nonconservative side of the Tarski boundary for ZF since it can prove $\mathsf{GRef}_{\mathsf{ZF}}$.

  1. $\mathsf{CT}_*[\mathsf{ZF}] := \mathsf{CT}^-[\mathsf{ZF}] + \mathsf{GRef}_{\mathsf{ZF}}$.
  2. $\mathsf{CT}_*[\mathsf{ZF}] + \mathsf{Th}(V) \in V$, where $\mathsf{Th}(V) \in V$ is the sentence expressesing:

$$\exists x (x = \{\sigma : \mathsf{T}(\sigma) \text{ and } \sigma \in \mathsf{Sent}_{\mathsf{ZF}}\}).$$

  3. $\mathsf{CT}_*[\mathsf{ZF}] + \forall x\ \exists y\ (y = \mathsf{T} \cap x)$.

- **Theorem.** *The above three theories are all subtheories of* $\mathsf{CT}_0[\mathsf{ZF}]$*; and their consistency can be verified in* $\mathsf{CT}_0[\mathsf{ZF}]$.
- **Theorem.** *The above three theories are listed in increasing order of deductive strength, as well as consistency strength.*

About $\mathrm{CT}_*[\mathrm{ZF}]$

- **Theorem.** GB_* *and* $\mathrm{CT}_*[\mathrm{ZF}]$ *are mutually* V*-interpretable*; *Consequently, they have the same* $\mathcal{L}_{\mathrm{ZF}}$*-consequences*.

  In contrast with the fact that $\mathrm{CT}_*[\mathrm{ZF}]$ proves that ZF has a model, we have;

- **Theorem.** $\mathrm{CT}_*[\mathrm{ZF}]$ *does not prove the existence of an* ω*-standard model of* ZF; *in particular the existence of a transitive model of* ZF *is not provable in* $\mathrm{CT}_*[\mathrm{ZF}]$.

- **Proof.** This follows from the following observations:

  1. If $\mathcal{M} \models \mathrm{ZF}$, and $\mathcal{M}$ is ω-standard, then $(\mathcal{M}, \mathrm{Th}(\mathcal{M}^+)) \models \mathrm{CT}_*[\mathrm{ZF}]$.
  2. ZF has an ω-model $\mathcal{M}$ that satisfies "ZF has no ω-model".

  Statement (1) is easy to see; (2) can be established using the abstract form of the second incompleteness theorem formulated in terms of Hilbert-Bernays-Löb provability conditions.

Glimpse of another many faces theorem

- **Theorem.** *The following theories have the same $\mathcal{L}_{\mathsf{ZF}}$-consequences as* $\mathsf{CT}_*[\mathsf{ZF}]$:
  1. $\mathsf{CT}^-[\mathsf{ZF}]$ + Int-Repl + "T is closed under first order proofs".
  2. $\mathsf{CT}^-[\mathsf{ZF}]$ + Int-Repl + "T contains all $\mathcal{L}^+_{\mathsf{ZF}}$-instances of theorems of first order logic".
  3. $\mathsf{CT}^-[\mathsf{ZF}]$ + Int-Repl + "T is closed under proofs of propositional logic".
  4. $\mathsf{CT}^-[\mathsf{ZF}]$ + Int-Repl + DC.
  5. $\mathsf{CT}^-[\mathsf{ZF}]$ + Int-Repl + DC-out.

About $\mathrm{CT}_*[\mathrm{ZF}] + \mathrm{Th}(V) \in V$

- **Theorem.** $\mathrm{CT}_*[\mathrm{ZF}] + \mathrm{Th}(V) \in V$ *does not prove that* ZF *has a model of the form* $(V_\alpha, \in)$.

  Proof. Let α be the first ordinal such that $(V_\alpha, \in) \models \mathrm{ZF}$, then $(V_\alpha, \in, \mathrm{Th}(\mathrm{V}_\alpha, \in)^+)$ satisfies:

  (1) $\mathrm{CT}_*[\mathrm{ZF}] + \mathrm{Th}(V) \in V$, and

  (2) There is no β such that $(V_\beta, \in)$ expands to $\mathrm{CT}_*[\mathrm{ZF}] + \mathrm{Th}(V) \in V$.

- **Theorem.** $\mathrm{CT}_*[\mathrm{ZF}] + \mathrm{Th}(V) \in V$ *proves that* ZF *a transitive model.*

- **Proof Outline**. It is a classical theorem (due to Jeff Paris) that every consistent extension of ZF has a model all of whose ordinals are pointwise definable. This theorem is provable in ZF itself.
  So within $\mathrm{CT}_*[\mathrm{ZF}] + \mathrm{Th}(V) \in V$ we can get hold of a model of $\mathrm{Th}(\mathcal{M})$ of ZF all of whose ordinals are pointwise definable.
  Then we can use internal replacement to show that $\mathcal{M}$ must be well-founded as viewed within $\mathrm{CT}_*[\mathrm{ZF}] + \mathrm{Th}(V) \in V$, and therefore $\mathcal{M}$ can be collapsed to a transitive model.

References for Lecture 3

E-1 A. Enayat, *Models of set Theory with definable ordinals*, **Archive for Mathematical Logic**, vol. 44 (2005), pp. 363-385.

E-2 A. Enayat, *The Mostowski Bridge*, **arXiv:2505.23998** (2025).

E-3 A. Enayat, *Tarskian truth theories over set theory*, in preparation.

EV A. Enayat and A. Visser, *Full satisfaction classes in a general setting*, privately circulated manuscript (2012).

F K. Fujimoto, *Classes and truths in set theory*, **Annals of Pure and Applied Logic**, vol. 163, (2012), pp. 1484-1523.

K S. Krajewski, *Nonstandard satisfaction classes*, in **Set Theory and Hierarchy Theory: A Memorial Tribute to Andrzej Mostowski** (edited by W. Marek et al.) Lecture Notes in Mathematics, vol. 537, Springer-Verlag, Berlin, 1976, pp. 121-144.

W B. Wcisło, *Truth and Collection*, **arXiv:2403.19367** (2024).

END OF LECTURE 3